Probability in the Engineering and Informational Sciences, page 1 of 19, 2017.
doi:10.1017/S0269964816000553

RARE EVENT ANALYSIS AND EFFICIENT SIMULATION FOR A MULTI-DIMENSIONAL RUIN PROBLEM

EWAN JACOV CAHEN
CWI, Amsterdam, the Netherlands,
E-mail: ewan.cahen@cwi.nl

MICHEL MANDJES
CWI, Amsterdam, the Netherlands,
University of Amsterdam, Amsterdam, the Netherlands
E-mail: m.r.h.mandjes@uva.nl

BERT ZWART
CWI, Amsterdam, the Netherlands
Eindhoven University of Technology, Eindhoven, the Netherlands
E-mail: bert.zwart@cwi.nl

This paper focuses on the evaluation of the probability that both components of a bivariate stochastic process ever simultaneously exceed some large level; a leading example is that of two Markov fluid queues driven by the same background process ever reaching the set $(u,\infty)\times(u,\infty)$, for $u>0$. Exact analysis being prohibitive, we resort to asymptotic techniques and efficient simulation, focusing on large values of u. The first contribution concerns various expressions for the decay rate of the probability of interest, which are valid under Gärtner–Ellis-type conditions. The second contribution is an importance-sampling-based rare-event simulation technique for the bivariate Markov modulated fluid model, which is capable of asymptotically efficiently estimating the probability of interest; the efficiency of this procedure is assessed in a series of numerical experiments.

Keywords: large deviations, importance sampling, rare-event simulation, ruin probability

1. INTRODUCTION

Let $((A_t,B_t))_{t\geq 0}$ be a bivariate stochastic process, with possibly dependent components. This paper focuses on techniques to quantify the so-called (bivariate) *ruin probability* over level u, denoted by α_u, being defined as the probability that this process will ever hit the set $S_u:=(u,\infty)\times(u,\infty)$, for some $u>0$. Note that even in the case that the two processes are independent, this is probability cannot be evaluated from the corresponding one-dimensional ruin probabilities, since both components have to be bigger than u at the *same time*.

 0269-9648/17 $25.00

A leading example of such a bivariate model is the *two-dimensional Markov modulated fluid model*, which can be described as follows. Let $(X_t)_{t\geq 0}$ be an irreducible Markov process, taking values on a finite state space. Whenever $X_t = i$, both A_t and B_t change at constant, possibly negative, rates r_i^A and r_i^B respectively so that

$$(A_t, B_t) = \left(\int_0^t r_{X_s}^A \, ds, \int_0^t r_{X_s}^B \, ds \right);$$

the two processes $(A_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ thus depend on each other as they react to the same realization of the *background process* $(X_t)_{t\geq 0}$. The process (A_t, B_t) could be used to model the evolution of two random quantities, which are driven by the same environment. Many examples can be thought of: the process can represent data buffers in a wireless network whose dynamics react to the same variations in the channel conditions, or multiple asset prices reacting to the same market fluctuations. The one-dimensional counterpart of this model is well understood. In particular, techniques have been developed to evaluate the ruin probabilities, by setting up a system of linear differential equations which can be solved by imposing the appropriate boundary conditions; see for example, Elwalid and Mitra [4]. Importantly, in the two-dimensional case these methods fail.

Since an exact analysis of α_u has been beyond reach so far, in this paper we turn to two approximative techniques that are intended to gain insight into the quantitative properties of α_u. The first technique is of an *asymptotic* nature: it characterizes the (essentially exponential) tail behavior of α_u for large u. The second approach is an efficient simulation technique based on importance sampling; it remedies the complication that straightforward, naïve simulation methods are typically slow due to the rarity of the event under consideration.

The research reported on in this paper is in the tradition of a series of papers on large deviations estimates and importance sampling for queues. For an introduction to importance sampling, we refer to for example, Rubino and Tuffin [12]. Importance sampling is a variance-reduction technique, which essentially amounts to sampling under another measure than the actual one, recovering unbiasedness by weighing the simulation data by appropriate likelihood ratios; the complication lies in the selection of the new measure, which should ideally be chosen such that the variance of the resulting estimator is minimized. Part of the paper relates to relatively general bivariate processes, and part to the specific case of bivariate Markov fluid. Asymptotics and efficient simulation for the one-dimensional model have been studied in detail; see for example, Mandjes and Ridder [11] and Rubino and Tuffin [12, Section 5.3.3]. We also mention Kesidis et al. [9], where to focus lies on the existence of so-called *effective bandwidths*; as it turns out, despite the fact that this work focuses on one-dimensional Markov fluid, results from this paper turn out to be useful in the context of our two-dimensional setup.

There is a vast literature that directly relates to the material presented in this paper; without aiming to give a complete overview, we mention a number of relevant contributions. In Glynn and Whitt [7], for a broad class of queues the exponential decay rate of the waiting time distribution is given; this result can be translated into the context of ruin probabilities. In Duffield and O'Connell [3], it is generalized to continuous time, as well as to non-linear scaling. Collamore [2] considers a related result for multi-dimensional discrete time Markov additive processes. In Kella [8], a fluid model is considered as well, but only for a Lévy input process (i.e., without Markov modulation); the main result is an expression for the Laplace–Stieltjes transform of the joint steady-state distribution.

This paper has two main results. The first one is Theorem 2.2, which gives multiple equivalent expressions for the *decay rate*

$$\lim_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t > 0 : A_t > u, B_t > u\right)$$

under mild assumptions. Importantly, these assumptions are met by the bivariate Markov fluid model. In the first part of the proof, we interpret a representation of the decay rate as the solution of a concave optimization problem with respect to several constraints. The use of this interpretation we believe is novel; for example, using this argumentation would have led to a considerably shorter proof in the one-dimensional case in Ganesh, O'Connell, and Wischik [6, Lemma 1.7]. A different expression for the decay rate is provided as well, with a proof that is split into two parts. First the lower bound is proven, which focuses on the largest contribution to α_u (in terms of a "dominant time scale"). For the upper bound, we determine the decay rate of the probability that the bivariate process hits some set T_u, which contains S_u; we let T_u be as small as possible so as to still obtain the correct decay rate. This approach is similar to the one used by Kosiński et al. [10].

The second main result concerns efficiency properties of an importance-sampling-based simulation scheme, which applies to the bivariate Markov fluid model only; it states that the underlying new measure is optimal (in a specific asymptotic sense). The new measure we propose can be regarded as the two-dimensional analogue of the change of measure that was used for the one-dimensional model in for example, Mandjes and Ridder [11] and Rubino and Tuffin [12, Section 5.3.3]. A complication is that the process (A_t, B_t) can attain values in the target set while the "embedded process" (recording values of (A_t, B_t) only at transition epochs of the background process) does not; we describe a technique to remedy this.

The rest of the paper is organized as follows. Section 2 contains the first main result, namely the decay rate for the general two-dimensional stochastic process under a Gärtner–Ellis (GE)-type condition. In Section 3, we specifically consider the bivariate Markov fluid model; we first present a number of results for this model, then we develop an efficient simulation algorithm, and finally we present a number of illustrative numerical examples. For readability, the proof of Theorem 3.1 is given in the Appendix. The paper concludes with Section 4, in which we discuss two natural extensions of the theory developed in Section 2, namely the extension of the theory to higher dimensions and the extension to bivariate processes for which the components may hit level u at *different* times.

2. LOGARITHMIC ASYMPTOTICS UNDER GÄRTNER-ELLIS CONDITIONS

Let $((A_t, B_t))_{t\geq 0}$ be a bivariate stochastic process on $\mathbb{R}^2$. We are interested in the probability α_u that the process will ever hit the set $S_u := (u,\infty)\times(u,\infty)$, for $u \gg 0$, when the average movement of the process is directed away from this set. More specifically, we wish to characterize the decay rate of this probability, that is,

$$\lim_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t : A_t > u, B_t > u\right). \tag{1}$$

We consider the situation that

$$\left(\lim_{t\to\infty} \frac{\mathbb{E}A_t}{t}, \lim_{t\to\infty} \frac{\mathbb{E}B_t}{t}\right) \notin [0,\infty)\times[0,\infty),$$

so that the event of interest is indeed rare. We will also assume that the process can reach the set with a positive probability. Note that we do not have to restrict the event of interest

to both components having to reach the *same* level, that is, the results developed in this paper can also handle the event

$$\{\exists t > 0 : A_t > u, B_t > cu\}$$

for any $c > 0$. This can be done by applying the analysis to the stochastic process $((A_t, B_t/c))_{t\geq 0}$. Hence, we will choose $c = 1$ in the remainder of this paper.

In order to be able to analyze the above decay rate, we now provide some results from large deviations theory, following the setup of Ganesh et al. [6]. We denote the *limiting cumulant generating function* of $((A_t, B_t))_{t\geq 0}$ by

$$M(\theta_1, \theta_2) := \lim_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(e^{\theta_1 A_t + \theta_2 B_t}\right). \tag{2}$$

A function $I : \mathbb{R}^d \to \mathbb{R}^*$ (where $\mathbb{R}^* := \mathbb{R} \cup \{\infty\}$) is a *rate function* if it is non-negative and if it is lower semi-continuous, that is, all level sets are closed. Furthermore, it is called a *good* rate function if in addition all level sets are compact. We say that $((A_t, B_t))_{t\geq 0}$ satisfies a *large deviations principle* in $\mathbb{R}^2$ with rate function $I : \mathbb{R}^2 \to \mathbb{R}^*$ if for any measurable set $F \subseteq \mathbb{R}^2$

$$-\inf_{x\in F^\circ} I(x) \leq \liminf_{t\to\infty} \frac{1}{t} \ln \mathbb{P}\left((A_t, B_t) \in F\right) \leq \limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{P}\left((A_t, B_t) \in F\right) \leq -\inf_{x\in \bar{F}} I(x),$$

where F° and $\bar{F}$ denote the interior and closure of F, respectively. For any function $f : \mathbb{R}^d \to \mathbb{R}^*$, we denote its *convex conjugate* by $f^*(x) := \sup_\theta \langle \theta, x\rangle - f(\theta)$. A function $f : \mathbb{R}^d \to \mathbb{R}^*$ is called *essentially smooth* if the interior of its effective domain (the set on which f is finite-valued) is non-empty, f is differentiable in the interior of its effective domain and f is steep, namely, for any sequence x_n, which converges to a boundary point of the effective domain, $\lim_{n\to\infty} |\nabla f(x_n)| = \infty$. The following well-known theorem will be used in the proof of our main result.

THEOREM 2.1 [Gärtner–Ellis, see Ganesh et al. [6, Theorem 2.11]]: *If* (2) *exists for all* θ_1, θ_2*, possibly taking value infinity, and if it is essentially smooth, lower semi-continuous and finite in a neighborhood of the origin, then the process* $((A_t/t, B_t/t))_{t\geq 0}$ *satisfies a large deviations principle in* $\mathbb{R}^2$ *with good convex rate function* M^**.*

Our main result gives the aforementioned decay rate in terms of the corresponding limiting cumulant generating function. Aside from assuming GE conditions, we also want to ensure that the continuous-time process is locally well-behaved so that we can apply the one-dimensional result from Duffield and O'Connell [3]. In order to make this precise, we define for $n \in \mathbb{N}$:

$$(A_n + B_n)^* := \sup_{0\leq r<1} (A_{n+r} + B_{n+r}).$$

THEOREM 2.2: *Let* $M(\cdot, \cdot)$ *satisfy the conditions of the GE theorem and let either*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{E}\left(e^{(\theta_1 A_n + \theta_2 B_n)^* - \theta_1 A_n - \theta_2 B_n}\right) = 0 \tag{3}$$

for all $\theta_1, \theta_2 > 0$*, or let* (3) *hold for some* $\theta_1, \theta_2 > 0$ *and let*

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\left((wA_n + (1-w)B_n)^* - wA_n - (1-w)B_n > xn\right) \leq -I(x, x) \tag{4}$$

hold for all $x > 0$ and all $w \in [0,1]$, with $I(\cdot,\cdot)$ as in (6). Then

$$\lim_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t > 0 : A_t > u, B_t > u\right) = -\inf_{x>0,y>0} \frac{I(x,y)}{\min(x,y)} \tag{5}$$

holds, where

$$I(x,y) := \sup_{\theta_1,\theta_2} \left(\theta_1 x + \theta_2 y - M(\theta_1,\theta_2)\right). \tag{6}$$

Furthermore,

$$\inf_{x>0,y>0} \frac{I(x,y)}{\min(x,y)} = \sup_{\theta_1\geq 0,\theta_2\geq 0: M(\theta_1,\theta_2)=0} (\theta_1+\theta_2). \tag{7}$$

PROOF: We begin with the proof of the latter statement, that is, (7). To this end, write

$$\inf_{x,y>0} \frac{I(x,y)}{\min(x,y)} = \inf_{x,y>0} \sup_{\theta_1,\theta_2} \frac{\theta_1 x + \theta_2 y}{\min(x,y)} - \frac{M(\theta_1,\theta_2)}{\min(x,y)}.$$

Setting $p := \min(x,y)$ we obtain

$$\inf_{x,y>0} \sup_{\theta_1,\theta_2} \frac{\theta_1 x + \theta_2 y}{\min(x,y)} - \frac{M(\theta_1,\theta_2)}{\min(x,y)} = \inf_{x,y,p>0,p=\min(x,y)} \sup_{\theta_1,\theta_2} \frac{\theta_1 x}{p} + \frac{\theta_2 y}{p} - \frac{M(\theta_1,\theta_2)}{p}.$$

We may now replace in the infimum $p = \min(x,y)$ by $p \leq \min(x,y)$, so that we find

$$\begin{aligned}&\inf_{x,y,p>0,p=\min(x,y)} \sup_{\theta_1,\theta_2} \frac{\theta_1 x}{p} + \frac{\theta_2 y}{p} - \frac{M(\theta_1,\theta_2)}{p}\\ &= \inf_{x,y,p>0,p\leq\min(x,y)} \sup_{\theta_1,\theta_2} \frac{\theta_1 x}{p} + \frac{\theta_2 y}{p} - \frac{M(\theta_1,\theta_2)}{p};\end{aligned}$$

this equality holds because

- The "$\geq$" part holds because the infimum on the right-hand side is taken over a larger set.
- The "$\leq$" part holds because the supremum is non-negative (choose $\theta_1 = \theta_2 = 0$), hence p will be taken as large as possible.

The next step is to replace the quantities x/p, y/p and $1/p$ by u, v, q respectively, which is allowed as long as we impose the restrictions $u, v \geq 1$ and $q > 0$ in the infimum. We thus obtain

$$\inf_{x,y,p>0,p\leq\min(x,y)} \sup_{\theta_1,\theta_2} \frac{\theta_1 x}{p} + \frac{\theta_2 y}{p} - \frac{M(\theta_1,\theta_2)}{p} = \inf_{u,v\geq 1,q>0} \sup_{\theta_1,\theta_2} \theta_1 u + \theta_2 v - qM(\theta_1,\theta_2).$$

We can now write $u\theta_1 = (1+a)\theta_1$ with $a \geq 0$, and similarly for θ_2 in order to obtain the alternative representation

$$\inf_{u,v\geq 1,q>0} \sup_{\theta_1,\theta_2} \theta_1 u + \theta_2 v - qM(\theta_1,\theta_2) = \inf_{a,b\geq 0,q>0} \sup_{\theta_1,\theta_2} (\theta_1+\theta_2) + a\theta_1 + b\theta_2 - qM(\theta_1,\theta_2).$$

The right-hand side of the previous display can now be seen as the Lagrangian dual of a concave optimization problem with respect to constraints as given in the infimum; it should

be borne in mind that the limiting cumulant generating function $M(\theta_1, \theta_2)$ is convex. As a consequence,

$$\inf_{a\geq 0, b\geq 0, q>0} \sup_{\theta_1,\theta_2} (\theta_1+\theta_2) + a\theta_1 + b\theta_2 - qM(\theta_1,\theta_2) = \sup_{\theta_1\geq 0,\theta_2\geq 0, M(\theta_1,\theta_2)\leq 0} \theta_1+\theta_2,$$

which establishes the proof of (7).

We will split the proof of the first statement, that is, (5), into a lower bound and an upper bound. We first give the lower bound. For all $s, u > 0$ we have the obvious bound

$$\mathbb{P}\left(\exists t>0 : A_t > u, B_t > u\right) = \mathbb{P}\left(\exists t>0 : A_{tu} > u, B_{tu} > u\right) \geq \mathbb{P}\left(A_{su} > u, B_{su} > u\right).$$

This means that also, for all $s > 0$,

$$\liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t>0 : A_{tu} > u, B_{tu} > u\right) \geq \liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(A_{su} > u, B_{su} > u\right).$$

As this inequality is uniform in $s > 0$, we can take the supremum on the right-hand side. We thus obtain, after rewriting:

$$\liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t : \frac{A_{tu}}{tu} > \frac{1}{t}, \frac{B_{tu}}{tu} > \frac{1}{t}\right) \geq \sup_{s>0} \liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\frac{A_{su}}{su} > \frac{1}{s}, \frac{B_{su}}{su} > \frac{1}{s}\right) \quad \textbf{(8)}$$

Furthermore, the GE theorem gives us the following inequality:

$$\begin{aligned}\liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\frac{A_{tu}}{tu} > \frac{1}{t}, \frac{B_{tu}}{tu} > \frac{1}{t}\right) &= t\left[\liminf_{u\to\infty} \frac{1}{tu} \ln \mathbb{P}\left(\frac{A_{tu}}{tu} > \frac{1}{t}, \frac{B_{tu}}{tu} > \frac{1}{t}\right)\right] \\ &\overset{\text{GE}}{\geq} -t \inf_{x>\frac{1}{t}, y>\frac{1}{t}} \left[\sup_{\theta_1,\theta_2} \theta_1 x + \theta_2 y - M(\theta_1,\theta_2)\right]. \quad \textbf{(9)}\end{aligned}$$

Upon combining (8) and (9), we thus conclude

$$\begin{aligned}&\liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t : A_t > u, B_t > u\right) = \liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\exists t : \frac{A_{tu}}{tu} > \frac{1}{t}, \frac{B_{tu}}{tu} > \frac{1}{t}\right) \\ &\overset{(8)}{\geq} \sup_{t>0} \liminf_{u\to\infty} \frac{1}{u} \ln \mathbb{P}\left(\frac{A_{tu}}{tu} > \frac{1}{t}, \frac{B_{tu}}{tu} > \frac{1}{t}\right) \overset{(9)}{\geq} \sup_{t>0} -t \inf_{x,y>\frac{1}{t}} \left[\sup_{\theta_1,\theta_2} \theta_1 x + \theta_2 y - M(\theta_1,\theta_2)\right] \\ &= \sup_{t>0} -t \inf_{x>\frac{1}{t}, y>\frac{1}{t}} I(x,y) = -\inf_{t>0} \inf_{x,y>\frac{1}{t}} tI(x,y) = -\inf_{x,y>0} \inf_{t>\max(\frac{1}{x},\frac{1}{y})} tI(x,y) = \\ &= -\inf_{x,y>0} \max\left(\frac{1}{x}, \frac{1}{y}\right) I(x,y) = -\inf_{x>0,y>0} \frac{I(x,y)}{\min(x,y)},\end{aligned}$$

which establishes the lower bound.

For the upper bound, we consider the probability of (A_t, B_t) reaching a set in which $(u,\infty)\times(u,\infty)$ is contained. We evidently have, for all "weights" $w \in [0,1]$,

$$\mathbb{P}\left(\exists t>0 : A_t > u, B_t > u\right) \leq \mathbb{P}\left(\exists t>0 : wA_t + (1-w)B_t > u\right);$$

for the moment we keep w fixed; later in the proof we minimize over w to identify the tightest upper bound. The crucial idea is that $wA_t + (1-w)B_t$ is now a *one-dimensional*

stochastic process, for which we can apply the result of Duffield and O'Connell [3, Corollary 2.3], so as to obtain

$$\limsup_{u\to\infty}\frac{1}{u}\ln\mathbb{P}\left(\exists t>0: A_t>u, B_t>u\right)\le\lim_{u\to\infty}\frac{1}{u}\ln\mathbb{P}\left(\exists t>0: wA_t+(1-w)B_t>u\right)$$
$$\le-\inf_{x>0}\frac{I_w(x)}{x}=-\theta_w^*,$$

where (i) the rate function $I_w(x)$ is defined by $\sup_\theta(\theta x-M_w(\theta))$; (ii) the limiting cumulant generating function $M_w(\theta)$ by

$$\lim_{t\to\infty}\frac{1}{t}\ln\mathbb{E}\left(e^{\theta wA_t+\theta(1-w)B_t}\right)=0,$$

and (iii) $\theta_w^*>0$ solves $M_w(\theta)=0$. Because the above upper bound on the decay rate holds for any $w\in[0,1]$, we can take the infimum with respect to w on both sides. We thus obtain

$$\limsup_{u\to\infty}\frac{1}{u}\ln\mathbb{P}\left(\exists t>0: A_t>u, B_t>u\right)\le-\sup_{w\in[0,1]}\theta_w^*.$$

Setting $\theta_1^*:=w\theta_w^*$ and $\theta_2^*:=(1-w)\theta_w^*$, we observe that $M(\theta_1^*,\theta_2^*)=0$. So then

$$\limsup_{u\to\infty}\frac{1}{u}\ln\mathbb{P}\left(\exists t>0: A_t>u, B_t>u\right)\le-\sup_{\theta_w^*>0,w\in[0,1]:M(w\theta_w^*,(1-w)\theta_w^*)=0}w\theta_w^*+(1-w)\theta_w^*$$
$$=-\sup_{\theta_1\ge0,\theta_2\ge0:M(\theta_1,\theta_2)=0}\theta_1+\theta_2,$$

which establishes the upper bound. ■

The first part of the proof uses a Lagrange-multiplier argument in order to show that the optimizing θ_1 and θ_2 are non-negative. This (seemingly novel) idea can be used more broadly; for instance in the proof of Ganesh et al. [6, Lemma 1.7]. In the proof of the upper bound, there is a one-to-one correspondence between w,θ_w^* and θ_1,θ_2, namely $\theta_1=w\theta_w^*$ and $\theta_2=(1-w)\theta_w^*$. A similar result is proven in Collamore [2] in a discrete time Markov additive setting, though the representation of the decay rate is different.

3. EFFICIENT ESTIMATION OF RUIN PROBABILITY IN BIVARIATE FLUID MODEL

In this section, as mentioned in the introduction, we let A_t and B_t represent two fluid processes, modulated by the *same* Markov process $(X_t)_{t\ge0}$ that attains values on a finite state space $\mathcal{N}$. We apply the theory of the previous section to set up an importance-sampling-based efficient simulation procedure for estimating

$$\alpha_u=\mathbb{P}\left(\exists t>0: A_t>u, B_t>u\right)=\mathbb{P}\left(\exists t>0:(A_t,B_t)\in S_u\right).$$

To make the model precise, let r^A and r^B be two vectors in $\mathbb{R}^{|\mathcal{N}|}$. Whenever $X_t=i$, the net input per time unit of the two components are r_i^A and r_i^B, respectively; note that these numbers are not necessarily positive. A compact representation is

$$\frac{dA_t}{dt}=r_i^A,\quad\frac{dB_t}{dt}=r_i^B\quad\text{if}\,X_t=i,\tag{10}$$

where we set $A_0=B_0=0$.

In the following subsection, we analyze the above model in greater detail, and conclude that it satisfies the conditions of Theorem 2.2. In Section 3.2 we construct a method in order to estimate the ruin probability efficiently and in Section 3.3 we give numerical examples (which also indicate the efficiency gain with respect to naïve simulation approaches).

3.1. Analysis of the Fluid Model

In order for the process to have a positive probability to hit the set S_u, some conditions have to be imposed on r^A and r^B. These conditions can be satisfied in essentially two ways. The first way is that there exists some state i such that $r_i^A > 0$ and $r_i^B > 0$. If such a state does not exist, we need two states i and j such that $r_i^A > 0, r_i^B \leq 0, r_j^A \leq 0, r_j^B > 0$ such that if the Markov process spends time in those states in a correct ratio, both A_t and B_t increase. A geometrical intuition for these conditions is given in Figure 1.

The following theorem characterizes these conditions, and shows that if S_u can be reached, it can do so by using, indeed, only at most two of the states of the modulating Markov process. Its (algebraic) proof is postponed to the appendix.

THEOREM 3.1: *The following three statements are equivalent:*

1. *The joint process can reach the set S_u, in that $\mathbb{P}(\exists t : A_t > u, B_t > u) > 0$.*
2. *There exist $i, j \in \mathcal{N}$ (possibly $i = j$) and $c, d \geq 0$ such that $cr_i^A + dr_j^A > 0$ and $cr_i^B + dr_j^B > 0$.*
3. *There exists $\vec{c} \in \mathbb{R}_+^{|\mathcal{N}|}$ such that $\langle \vec{c}, r^A \rangle > 0$ and $\langle \vec{c}, r^B \rangle > 0$, where $\langle \cdot, \cdot \rangle$ denotes the inner product.*

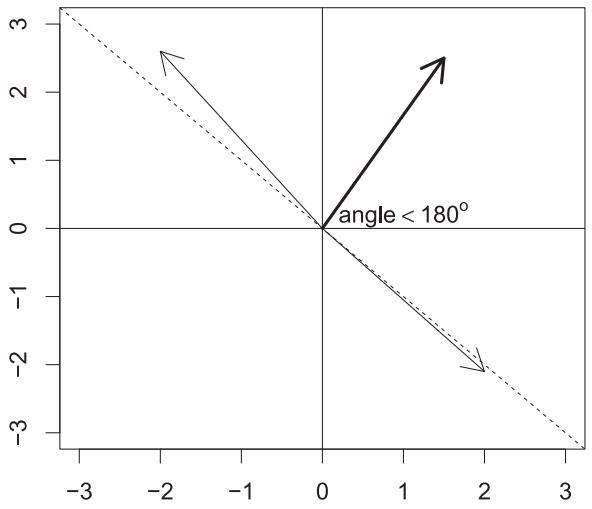

FIGURE **1.** This figure illustrates the conditions under which the process can hit the set $S_u = (u, \infty) \times (u, \infty)$. The arrows represent the direction the process is going when X_t remains in some state for one time unit. The first possibility is that there is an arrow in the upper-right quadrant, like the bold arrow. The other possibility is that there are two arrows in the upper-left and lower-right quadrant respectively, of which at least one is above the dashed line and the other makes an angle less than 180° with the first one. An example for this possibility is given by the two non-bold arrows.

In order to avoid trivialities, we also assume that both components have a negative drift, that is, $\langle r^A, \pi\rangle < 0$ and $\langle r^B, \pi\rangle < 0$, where π denotes the equilibrium distribution of X.

It is not clear yet that this process satisfies the conditions of the GE theorem. The following theorem, which is a generalization of Kesidis et al. [9, page 5], shows that it indeed does. Note also that condition (3) is satisfied trivially, since $(\theta_1 A_n + \theta_2 B_n)^* - \theta_1 A_n - \theta_2 B_n$ is uniformly bounded in n. We can thus apply Theorem 2.2 to this model.

THEOREM 3.2: *The value of $M(\theta_1, \theta_2)$ is equal to the largest real eigenvalue of the matrix*

$$Q + \theta_1 R^A + \theta_2 R^B,$$

where R^A and R^B are diagonal matrices with the rate vectors r^A and r^B, respectively, on their diagonals. Furthermore, $M(\cdot,\cdot)$ is differentiable.

PROOF: We will first derive an expression for $f(t) := \mathbb{E}\left(\exp(\theta_1 A_t + \theta_2 B_t)\right) = \sum_i f_i(t)$, where

$$f_i(t) := \mathbb{E}\left(e^{\theta_1 A_t + \theta_2 B_t} \mathbb{1}\left(X(t) = i\right)\right).$$

Relying on standard "Markovian reasoning]", as $\Delta \downarrow 0$,

$$f_i(t) = \sum_{k\neq i} f_k(t-\Delta) q_{ki} \Delta e^{\theta_1 r^A_{ki}\Delta} e^{\theta_2 r^B_{ki}\Delta} + f_i(t-\Delta)(1 - q_i\Delta) e^{\theta_1 r^A_i \Delta} e^{\theta_2 r^B_i \Delta} + o(\Delta).$$

By writing exponentials as power series we straightforwardly obtain:

$$\begin{aligned} f_i(t) = &\sum_{k\neq i} f_k(t-\Delta) q_{ki}\Delta(1+\theta_1 r^A_{ki}\Delta)(1+\theta_2 r^B_{ki}\Delta) \\ &+ f_i(t-\Delta)(1-q_i\Delta)(1+\theta_1 r^A_i\Delta)(1+\theta_2 r^B_i\Delta) + o(\Delta), \end{aligned}$$

which simplifies to

$$f_i(t) = \sum_{k\neq i} f_k(t-\Delta) q_{ki}\Delta + f_i(t-\Delta)(1 - q_i\Delta + \theta_1 r^A_i\Delta + \theta_2 r^B_i\Delta) + o(\Delta).$$

Rearranging and dividing by Δ gives

$$\frac{f_i(t) - f_i(t-\Delta)}{\Delta} = \sum_{k\neq i} f_k(t-\Delta) q_{ki} + f_i(t-\Delta)(-q_i + \theta_1 r^A_i + \theta_2 r^B_i) + o(1).$$

Now letting $\Delta \downarrow 0$ and realizing that $q_i := -q_{ii} = \sum_{k\neq i} q_{ki}$,

$$f_i'(t) = \sum_k f_k(t) q_{ki} + f_i(t)(\theta_1 r^A_i + \theta_2 r^B_i),$$

which is in matrix–vector notation equivalent to $f'(t) = (Q^T + \theta_1 R^A + \theta_2 R^B) f(t)$. This system of linear differential equations is solved by $f(t) = \exp((Q^T + \theta_1 R^A + \theta_2 R^B)t)\, f(0)$. Along the lines of Kesidis et al. [9, page 5], the first result now follows.

Note that $\exp(Q^T + \theta_1 R^A + \theta_2 R^B)$ has positive entries only. This can be seen by choosing some $a > 0$ large enough such that $Q + \theta_1 R^A + \theta_2 R^B + aI \geq 0$ and hence

$$e^{Q+\theta_1 R^A + \theta_2 R^B} = e^{Q+\theta_1 R^A + \theta_2 R^B + aI - aI} = e^{-a} e^{Q+\theta_1 R^A + \theta_2 R^B + aI} > 0.$$

The last strict inequality holds because Q is irreducible, and thus so is $Q + \theta_1 R^A + \theta_2 R^B + aI$, and then according to Seneta [13, Lemma 1.3] some power of this matrix is positive. Note

that each entry of $\exp(Q^T + \theta_1 R^A + \theta_2 R^B)$ can be written as a power series in the variables θ_1 and θ_2, so each entry of $\exp(Q^T + \theta_1 R^A + \theta_2 R^B)$ is infinitely many times differentiable with respect to these variables. It thus follows from Biggins and Sani [1, Proposition 1] that M is differentiable. ■

3.2. Simulation: Construction of an Efficient Method

As discussed in the introduction, we propose to use importance sampling in order to efficiently estimate α_u. In this section, we identify an appropriate new measure $\mathbb{Q}$, adopting an approach similar to that used in the one-dimensional case; see for example, Mandjes and Ridder [11]. There it is argued that the decay rate can be found by solving the eigensystem $-\theta^* R^A x = Qx$, that is, by calculating the eigenvectors and eigenvalues of $(R^A)^{-1}Q$, where the non-negative eigenvector that corresponds to the largest negative eigenvalue, is used in the change of measure; such eigenvector/eigenvalue pair exists due to "Perron–Frobenius". For the two-dimensional model, where the analogous eigensystem is $-(\theta_1 R^a + \theta_2 R^b)x = Qx$, we can not use this method any more, since left-hand side of the eigensystem can not be inverted without knowing θ_1 and θ_2 beforehand. However, we can combine Theorems 2.2 and 3.2 in order to find θ_1^* and θ_2^* numerically, by using binary search on the value of $\theta_1 + \theta_2$. Once we have found θ_1^* and θ_2^*, we can use the eigensystem to calculate the appropriate eigenvector.

THEOREM 3.3: *For all θ_1, θ_2 such that $M(\theta_1, \theta_2) = 0$, there exists $x \in \mathbb{R}_+^{|\mathcal{N}|}$ such that $-(\theta_1 R^A + \theta_2 R^B)x = Qx$.*

PROOF: Since the value of $M(\theta_1, \theta_2)$ is equal to the largest real eigenvalue of $Q + \theta_1 R^A + \theta_2 R^B$, it follows from Seneta [13, Theorem 2.5(e) (pp. 40, 41)] that

$$q_i > \theta_1 r_i^A + \theta_2 r_i^B$$

for all i. We need this for the existence of specific moment generating functions below. Let

$$x_{ii} := \mathbb{E}\left(\exp(\theta_1 \tilde{A}_{ii} + \theta_2 \tilde{B}_{ii})\right),$$

where $\tilde{A}_{ii}$ ($\tilde{B}_{ii}$) denotes the net amount of fluid generated by A_t (B_t) between two consecutive visits of the Markov process to state i. Likewise, we let

$$x_{ij} := \mathbb{E}\left(e^{\theta_1 \hat{A}_{ij} + \theta_2 \hat{B}_{ij}}\right),$$

$i \neq j$, where $\hat{A}_{ij}$ ($\hat{B}_{ij}$) denotes the net amount of fluid generated by A_t (B_t) between a visit of the Markov process to state i and the next visit to state $j \neq i$. We can then write, by conditioning on the first state the Markov process visits,

$$x_{ii} \equiv x_{ii}(\theta_1, \theta_2) = \sum_{j \neq i} \frac{\lambda_{ij}}{\lambda_i} \frac{\lambda_i}{\lambda_i - \theta_1 r_i^A - \theta_2 r_i^B} x_{ji}.$$

Using the reasoning of Mandjes and Ridder [11], solutions of $M(\theta_1, \theta_2) = 0$ also solve $x_{ii}(\theta_1, \theta_2) = 1$. Canceling the q_i and multiplying by the denominator of the right-hand side

gives us

$$(\lambda_i - \theta_1 r_i^A - \theta_2 r_i^B)x_{ii} = \sum_{j\neq i} q_{ij}x_{ji},$$

which is equivalent to

$$(-\theta_1 r_i^A - \theta_2 r_i^B)x_{ii} = \sum_{j} q_{ij}x_{ji}.$$

In the same way we can also write, for $j \neq i$,

$$x_{ji} = \frac{q_{ji}}{q_j - \theta_1 r_j^A - \theta_2 r_j^B} + \sum_{d\neq j, d\neq i} \frac{q_{jd}}{q_j - \theta_1 r_j^A - \theta_2 r_j^B} x_{di},$$

which can be rewritten as

$$(-\theta_1 r_j^A - \theta_2 r_j^B)x_{ji} = q_{ji} + \sum_{d\neq i} q_{jd}x_{di}.$$

As we have that $x_{ii} = 1$, we obtain, for all i, j that

$$(-\theta_1 r_j^A - \theta_2 r_j^B)x_{ji} = \sum_{d} q_{jd}x_{di}.$$

For fixed i, this can be rewritten as $-(\theta_1 R^A + \theta_2 R^B)x_i = Qx_i$, with $x_i := (x_{ji})_j$. ■

The new measure $\mathbb{Q}$ under which we sample $(X_t)_{t\geq 0}$ is then constructed as follows. Let θ_1^* and θ_2^* be the optimizing values resulting from Theorem 2.2, and let x be the corresponding eigenvector as given in Theorem 3.2. We replace $Q = (\lambda_{ij})_{i,j}$ with $\tilde{Q} = (\tilde{q}_{ij})_{i,j}$, where $\tilde{q}_{ij} := q_{ij}x_j/x_i$ for $i \neq j$ and $\tilde{q}_i := -\tilde{q}_{ii} = q_i - r_i^A\theta_1^* - r_i^B\theta_2^*$.

Our objective is to show that this is indeed a good change of measure, in the sense that it is *asymptotically optimal* (see Rubino and Tuffin [12, Definition 1, pp. 89, 90]). To this end, let L be the likelihood ratio of a path generated under the change of measure for which both $A_t > u, B_t > u$ for some $t > 0$. Denote by J_m the state of the Markov process after the mth jump and let T_m denote the time spent there. Furthermore, let

$$N := \inf\{n \in \mathbb{N} : \exists t \leq t_n \text{such that} A_t > u \text{and} B_t > u\},$$

that is, N is the smallest number of jumps until there was some t such that both components were bigger than u at time t. As pointed out in Rubino and Tuffin [12], the likelihood ratio then reads

$$L = \frac{\pi_{J_0}}{\varrho_{J_0}} \frac{q_{J_0J_1}}{\tilde{q}_{J_0J_1}} \cdots \frac{q_{J_{N-1}J_N}}{\tilde{q}_{J_{N-1}J_N}} \cdot \frac{q_{J_N}}{\tilde{q}_{J_N}} \cdot \exp\left(-\sum_{m=0}^{N}(q_{J_m} - \tilde{q}_{J_m})T_m\right), \tag{11}$$

with ϱ the invariant distribution of X_t under $\mathbb{Q}$.

We first point out a "naïve" implementation, which we denote by $\mathbb{Q}_1$. Start with some initial state X_0, sampled according to ϱ; say we draw j_0. Then sample, according to the corresponding exponential distribution (i.e., with parameter $\tilde{q}_{j_0}$), some time $t_0 > 0$ for which the Markov process remains in this state. We can then update the likelihood and calculate A_{t_0} and B_{t_0}. If both are bigger then u we stop; else we sample the next state, say j_1, using the probabilities $\tilde{q}_{j_0,k}/\tilde{q}_{j_0}$ for $k \neq j_0$. We continue with this procedure until both $A_{t_N} > u$ and $B_{t_N} > u$.

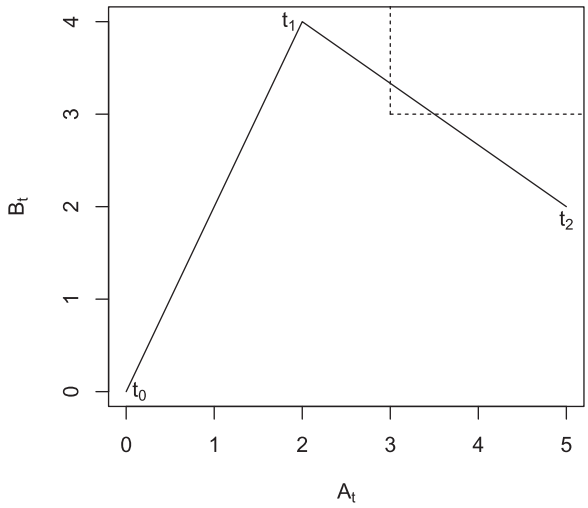

FIGURE **2.** The process hits the desired set, bounded by the dashed lines, between t_1 and t_2, but is not in the set at t_1 or t_2.

There is a complication, however. With the above procedure we only check at transition epochs of the modulating Markov process whether or not S_u has been reached. However, this poses a problem which does not occur in the one-dimensional process: it could happen that at two consecutive transition epochs the process did not reach the desired set, but at some time epoch *in between* these jumps, it did. This scenario is illustrated in Figure 2.

As a result, using this procedure we do not estimate α_u, but rather, with $U_n := T_0 + \ldots + T_n$,

$$\bar{\alpha}_u := \mathbb{P}\left(\exists n \in \mathbb{N} : A_{U_n} > u, B_{U_n} > u\right).$$

Clearly $\bar{\alpha}_u < \alpha_u$, creating a bias.

Whether or not the scenario of Figure 2 has occurred can, however, easily be checked from subsequent pairs of the form $(A_{t_{N-1}}, B_{t_{N-1}})$ and (A_{t_N}, B_{t_N}). If this happens, we propose to replace the factor

$$\frac{q_{J_N}}{\tilde{q}_{J_N}} \cdot \exp\left(-(q_{J_N} - \tilde{q}_{J_N})T_N\right) \tag{12}$$

in (11) by

$$\exp\left(-(q_{J_N} - \tilde{q}_{J_N})\tau\right), \tag{13}$$

where τ is the length of the interval between t_{N-1} and the first time epoch at which the process hit the desired set. The theorem below states that this adapted version of the naïve implementation, denoted by $\mathbb{Q}_2$, estimates α_u in an unbiased and asymptotically optimal way.

THEOREM 3.4: *The implementation $\mathbb{Q}_1$ yields an unbiased, asymptotically optimal estimate of $\bar{\alpha}_u$. The implementation $\mathbb{Q}_2$ yields an unbiased, asymptotically optimal estimate of α_u.*

PROOF: We start by proving the claim regarding $\mathbb{Q}_1$. Directly from the definition of the new rates $\tilde{q}_{ij}$ and the stopping time T_N,

$$\begin{aligned} L &= \frac{\pi_{J_0}}{\varrho_{J_0}} \frac{x_{J_0}}{x_{J_N}} \frac{q_{J_N}}{\tilde{q}_{J_N}} \cdot \exp\left(- \sum_{m=0}^{N} (q_{J_m} - \tilde{q}_{J_m}) T_m \right) \\ &= \frac{\pi_{J_0}}{\varrho_{J_0}} \frac{x_{J_0}}{x_{J_N}} \frac{q_{J_N}}{\tilde{q}_{J_N}} \cdot \exp\left(- \sum_{m=0}^{N} (r^A_{J_m} \theta^*_1 + r^B_{J_m} \theta^*_2) T_m \right) \leq k \cdot \exp\left(-(\theta^*_1 + \theta^*_2)u\right) \end{aligned}$$

with

$$k := \max_{i,j} \frac{\pi_i}{\varrho_i} \frac{x_i}{x_j} \frac{q_j}{\tilde{q}_j};$$

realize that

$$\sum_{m=0}^{N} r^A_{J_m} T_m > u, \quad \sum_{m=0}^{N} r^B_{J_m} T_m > u.$$

From this upper bound on the likelihood it follows that

$$\begin{aligned} \lim_{u\to\infty} \frac{1}{u} \ln \mathbb{E}_{\mathbb{Q}_1} \left(L^2 \mathbb{1} \left(\exists n \in \mathbb{N} : A_{U_n} > u, B_{U_n} > u \right) \right) &\leq \lim_{u\to\infty} \frac{1}{u} \ln \left[k \cdot e^{-2(\theta^*_1 + \theta^*_2)u} \right] \\ &= -2(\theta^*_1 + \theta^*_2). \end{aligned}$$

From this and Theorem 2.2, asymptotic optimality follows.

Regarding implementation $\mathbb{Q}_2$, realize that, because of the definition of τ,

$$\sum_{m=0}^{N-1} r^A_{J_m} T_m + r^A_{J_N} \tau > u, \quad \sum_{m=0}^{N-1} r^B_{J_m} T_m + r^B_{J_N} \tau > u.$$

In this case,

$$\begin{aligned} L &= \frac{\pi_{J_0}}{\varrho_{J_0}} \frac{x_{J_0}}{x_{J_N}} \frac{q_{J_N}}{\tilde{q}_{J_N}} \times \exp\left(-\theta^*_1 \left(\sum_{m=0}^{N-1} r^A_{J_m} T_m + r^A_{J_N} \tau \right) \right) \\ &\quad \times \exp\left(-\theta^*_2 \left(\sum_{m=0}^{N-1} r^B_{J_m} T_m + r^B_{J_N} \tau) \right) \right) \\ &\leq k \cdot \exp\left(-(\theta^*_1 + \theta^*_2)u\right). \end{aligned}$$

Asymptotic optimality follows as before. To show that $\mathbb{Q}_2$ indeed yields an unbiased estimate of α_u, we need to show that (12) and (13) have the same expectation under $\mathbb{Q}_2$ whenever $T_N > \tau$. Note that

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}_2} \left(e^{-(\lambda_{J_N} - \mu_{J_N})\tau} \mathbb{1} \left(T_N > \tau \right) \right) &= e^{-(\lambda_{J_N} - \mu_{J_N})\tau} \mathbb{P}_{\mathbb{Q}} \left(T_N > \tau \right) \\ &= e^{-(\lambda_{J_N} - \mu_{J_N})\tau} e^{-\mu_{J_N}\tau} = e^{-\lambda_{J_N}\tau}, \end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_2}\left(\frac{\lambda_{J_N}}{\mu_{J_N}}\cdot e^{-(\lambda_{J_N}-\mu_{J_N})T_N}\mathbb{1}\left(T_N>\tau\right)\right)\\
&=\frac{\lambda_{J_N}}{\mu_{J_N}}\mathbb{E}_{\mathbb{Q}_2}\left(e^{-(\lambda_{J_N}-\mu_{J_N})T_N}|\mathbb{1}\left(T_N>\tau\right)\right)\mathbb{P}_{\mathbb{Q}_2}\left(\mathbb{1}\left(T_N>\tau\right)\right)\\
&=\frac{\lambda_{J_N}}{\mu_{J_N}}e^{-(\lambda_{J_N}-\mu_{J_N})\tau}\mathbb{E}_{\mathbb{Q}_2}\left(e^{-(\lambda_{J_N}-\mu_{J_N})(T_N-\tau)}|\mathbb{1}\left(T_N>\tau\right)\right)\mathbb{P}_{\mathbb{Q}_2}\left(\mathbb{1}\left(T_N>\tau\right)\right)\\
&=\frac{\lambda_{J_N}}{\mu_{J_N}}e^{-(\lambda_{J_N}-\mu_{J_N})\tau}\mathbb{E}_{\mathbb{Q}_2}\left(e^{-(\lambda_{J_N}-\mu_{J_N})T_N}\right)\mathbb{P}_{\mathbb{Q}_2}\left(\mathbb{1}\left(T_N>\tau\right)\right)\\
&=\frac{\lambda_{J_N}}{\mu_{J_N}}e^{-(\lambda_{J_N}-\mu_{J_N})\tau}\frac{\mu_{J_N}}{\mu_{J_N}-(\lambda_{J_N}-\mu_{J_N})}e^{-\mu_{J_N}\tau}=e^{-\lambda_{J_N}\tau},
\end{aligned}$$

■

3.3. Numerical Results

We now consider some numerical examples. For the first example, we consider 20 on–off processes feeding into a two-dimensional reservoir.

Each on–off process generates, while on, traffic at constant rate 3 (4) into the first (second) reservoir. The first reservoir has a constant leak rate of 30.5, where for the second reservoir this is equal to 47.5. Note that the reservoirs have an equal net input when 17 of the sources are turned on. When a source is off, it will turn on at rate 2, while a working source turns off at rate 3. Since all sources behave identically, it suffices to take as state space $\mathcal{N}=\{0,\ldots,20\}$, where $X_t=i$ means that at time t there are i sources turned on. The rate matrix Q is then

$$Q=\begin{array}{c} \\ 0\\1\\2\\3\\ \vdots\\19\\20\end{array}\begin{array}{c}\begin{array}{ccccccccc}0&1&2&3&4&\cdots&18&19&20\end{array}\\ \begin{pmatrix}-40&40&&&&&&&\\3&-41&38&&&&&&\\&6&-42&36&&&&&\\&&9&-43&34&&&&\\&&&&&\ddots&&&\\&&&&&&57&-59&2\\&&&&&&&60&-60\end{pmatrix}\end{array},$$

and r^A and r^B are

$$\begin{aligned}
r^A&=\begin{pmatrix}-30.5&-27.5&-24.5&\cdots&26.5&29.5\end{pmatrix},\\
r^B&=\begin{pmatrix}-47.5&-44.5&-40.5&\cdots&28.5&32.5\end{pmatrix}.
\end{aligned}$$

The results can be found in Table 1.

The second example has the same structure as the first example. We now consider five on–off sources. The first reservoir has a constant service rate of 8.5, where for the second reservoir this is equal to 12.5. The other numbers are the same as for the first example. Note that the reservoirs have an equal net input when 4 of the sources are turned on. A key difference is that under the change of measure, in this example both queues have the same drift, whereas is the first example A_t has a higher drift than B_t. Informally this means that for the second example the two processes reach level u roughly simultaneously, while for the

TABLE 1. Simulation results of the first example. We denote by $\hat{\alpha}$ the estimated probabilities for both importance sampling and Monte Carlo sampling. Furthermore, we denote by θ^* the value of (7). This table also shows the number of samples needed to get a 95% confidence interval with 10% precision. The missing values took more than 100,000 samples in order to give the desired precision. In this case $\theta_1^* = 0, \theta_2^* = 1.004$.

	Importance sampling			Monte Carlo sampling	
u	$\hat{\alpha}$	$\hat{\alpha}e^{\theta^* u}$	# Samples	$\hat{\alpha}$	# Samples
0.25	2.95×10^{-2}	3.79×10^{-2}	4,884	2.78×10^{-2}	13,457
0.50	1.76×10^{-2}	2.90×10^{-2}	5,667	1.70×10^{-2}	22,212
1.00	7.97×10^{-3}	2.18×10^{-2}	6,846	8.22×10^{-3}	46,369
1.50	4.94×10^{-3}	2.23×10^{-2}	10,987	4.78×10^{-3}	80,170
2.00	2.24×10^{-3}	1.66×10^{-2}	5,113	–	–
3.00	7.61×10^{-4}	1.55×10^{-2}	8,244	–	–
4.00	2.83×10^{-4}	1.57×10^{-2}	14,962	–	–
5.00	9.01×10^{-5}	1.36×10^{-2}	8,493	–	–
6.00	3.42×10^{-5}	1.42×10^{-2}	9,893	–	–
7.00	1.14×10^{-5}	1.28×10^{-2}	9,886	–	–
8.00	4.63×10^{-6}	1.43×10^{-2}	10,354	–	–
9.00	1.66×10^{-6}	1.39×10^{-2}	12,600	–	–
10.00	5.87×10^{-7}	1.35×10^{-2}	12,777	–	–

first example the joint process will hit the set S_u when B_t does. The results can be found in Table 2. In both examples, the number of runs needed when using importance sampling is significantly lower than when ordinary Monte Carlo sampling is used. Furthermore, when using Monte Carlo sampling, the number of runs needed increases rapidly as a function of u, while there there is only a slight increase when using importance sampling.

In our next example the two processes are positively correlated. We let r^A and r^B be

$$r^A = \begin{pmatrix} -8 & -6 & 8 & 2 \end{pmatrix}, \quad r^B = \begin{pmatrix} -5 & -9 & 3 & 9 \end{pmatrix}.$$

We let Q contain only ones off the diagonal, so that all values on the diagonal are -3. We want to focus on the quality of the importance sampling results, so we did not do Monte Carlo sampling. The results can be found in Table 3.

It can be seen that the logarithmic result is very accurate as u grows large, as the fraction of the logarithmic result and θ^* tends to unity. However, from the next column we can see that by using the logarithmic asymptotics we lose some information, as the values in that column do not tend to a constant.

In the fourth example, we let the two processes be *negatively* correlated. We let r^A and r^B be

$$r^A = \begin{pmatrix} -8 & -6 & 8 & 2 \end{pmatrix}, \quad r^B = \begin{pmatrix} 3 & 9 & -5 & -9 \end{pmatrix}.$$

Note that r^A is the same as in the fourth example, and that r^B is a permutation of r^B as in the third example. The results can be found in Table 4.

It can be seen that the hitting probability decreases much faster than in the previous example. This can be explained by the fact that the Markov process in the previous example

TABLE 2. Simulation results of the second example. We denote by $\hat{\alpha}$ the estimated probabilities for both importance sampling and Monte Carlo sampling. Furthermore, we denote by θ^* the value of (7). This table also shows the number of samples needed to get a 95% confidence interval with 10% precision. The missing values took more than 100,000 samples in order to give the desired precision. In this case, $\theta_1^* = 0.128, \theta_2^* = 1.072$.

	Importance sampling			Monte Carlo sampling	
u	$\hat{\alpha}$	$\hat{\alpha}e^{\theta^* u}$	# Samples	$\hat{\alpha}$	# Samples
0.25	1.37×10^{-1}	1.85×10^{-1}	685	1.38×10^{-1}	2,407
0.50	9.91×10^{-2}	1.81×10^{-1}	794	1.03×10^{-1}	3,339
1.00	4.85×10^{-2}	1.61×10^{-1}	770	4.88×10^{-2}	7,507
1.50	2.49×10^{-2}	1.51×10^{-1}	825	2.69×10^{-2}	13,908
2.00	1.45×10^{-2}	1.60×10^{-1}	1,001	1.14×10^{-2}	33,320
3.00	4.08×10^{-3}	1.50×10^{-1}	856	4.04×10^{-3}	94,731
4.00	1.21×10^{-3}	1.47×10^{-1}	855	–	–
5.00	3.51×10^{-4}	1.42×10^{-1}	723	–	–
6.00	1.04×10^{-4}	1.40×10^{-1}	992	–	–
7.00	3.30×10^{-5}	1.47×10^{-1}	1,517	–	–
8.00	1.02×10^{-5}	1.52×10^{-1}	932	–	–
9.00	2.71×10^{-6}	1.34×10^{-1}	1,035	–	–
10.00	7.68×10^{-7}	1.26×10^{-1}	1,075	–	–

TABLE 3. Simulation results of the third example. We denote by $\hat{\alpha}$ the estimated probabilities. Furthermore, we denote by θ^* the value of (7). This table also shows the number of samples needed to get a 95% confidence interval with 10% precision. In this case $\theta_1^* = 0.094, \theta_2^* = 0.001$.

	Importance sampling			
u	$\hat{\alpha}$	$(-1/u\theta^*)\ln(\hat{\alpha})$	$\hat{\alpha}e^{\theta^* u}$	# Samples
1	6.37×10^{-1}	4.748	7.00×10^{-1}	73
2	6.14×10^{-1}	2.572	7.42×10^{-1}	40
3	5.23×10^{-1}	2.273	6.96×10^{-1}	67
4	4.46×10^{-1}	2.125	6.52×10^{-1}	84
5	4.72×10^{-1}	1.580	7.59×10^{-1}	54
6	3.62×10^{-1}	1.784	6.40×10^{-1}	96
7	3.28×10^{-1}	1.677	6.37×10^{-1}	100
8	2.98×10^{-1}	1.592	6.38×10^{-1}	89
9	3.11×10^{-1}	1.366	7.31×10^{-1}	70
10	2.58×10^{-1}	1.428	6.66×10^{-1}	80
20	8.60×10^{-2}	1.291	5.75×10^{-1}	129
50	4.80×10^{-3}	1.124	5.54×10^{-1}	181
100	3.39×10^{-5}	1.084	4.52×10^{-1}	260
250	2.00×10^{-11}	1.038	4.09×10^{-1}	344

TABLE 4. Simulation results of the fourth example. We denote by $\hat{\alpha}$ the estimated probabilities. Furthermore, we denote by θ^* the value of (7). This table also shows the number of samples needed to get a 95% confidence interval with 10% precision. In this case $\theta_1^* = 0.344, \theta_2^* = 0.325$.

	Importance sampling			
u	$\hat{\alpha}$	$(-1/u\theta^*)\ln(\hat{\alpha})$	$\hat{\alpha}e^{\theta^* u}$	# Samples
1	1.34×10^{-1}	3.002	2.62×10^{-1}	657
2	6.05×10^{-2}	2.098	2.30×10^{-1}	766
3	2.95×10^{-2}	1.757	2.19×10^{-1}	1,179
4	1.32×10^{-2}	1.619	1.91×10^{-1}	1,180
5	6.60×10^{-3}	1.502	1.87×10^{-1}	1,531
6	3.26×10^{-3}	1.428	1.80×10^{-1}	1,881
7	1.59×10^{-3}	1.377	1.72×10^{-1}	1,858
8	7.60×10^{-4}	1.343	1.60×10^{-1}	5,005
9	3.84×10^{-4}	1.307	1.58×10^{-1}	3,319
10	1.79×10^{-4}	1.291	1.43×10^{-1}	4,241
20	1.77×10^{-7}	1.163	1.13×10^{-1}	8,366
50	2.16×10^{-16}	1.079	7.10×10^{-2}	6,000
100	5.11×10^{-31}	1.043	5.54×10^{-2}	26,001
250	8.85×10^{-75}	1.020	3.43×10^{-2}	22,857

just needs to spend enough time in one of the to up-states for long enough, while in this example the process needs to visit *two* states long enough, and also has to do this in a correct ratio. Note that in this example it will happen frequently that the process will hit the set $(u,\infty)\times(u,\infty)$ in the way as depicted in Figure 2.

4. OUTLOOK

There are some natural extensions of the theory developed in Section 2 that can be investigated, of which we briefly comment on two. The first one is whether the two-dimensional theory can be extended to arbitrary higher dimensions. At the expense of introducing additional notation and loss of transparency, Theorem 2.2 can indeed be extended to a multi-dimensional result. For proving (5), the proof of the lower bound holds true trivially in multiple dimensions. In proving the upper bound, instead of using a single weight $w \in [0,1]$, one has to consider non-negative weights $w_1, w_2, \ldots, w_d$ that add up to unity. The proof of (7) does not use any specific two-dimensional argument, so is also valid for multiple dimensions.

The second extension considers a bivariate process for which the two components have to be larger than u but can do so at *different* times. Of course, this probability is greater than the probability of the event which is considered in this paper. A significant difference also occurs in simulating, because the simulation can now be split in two parts. First, one simulates under a change-of-measure until one of the components has hit level u. After this occurred, one can use a *different* change of measure until the other component hits level u; under this different change of measure, the first component to have hit level u may now also

have a negative drift, since it has already hit level u. The efficiency of this algorithm is not obvious, and is currently under investigation.

Acknowledgments

The research of E. J. Cahen has been funded by an NWO grant, under number 613.001.352

References

1. Biggins, J. & Sani, A., (2004). *Extended Perron–Frobenius results*. University of Sheffield, Department of Probability and Statistics.
2. Collamore, J. (1996). Hitting probabilities and large deviations. *The Annals of Probability*, 24(4): 2065–2078.
3. Duffield, N. & O'Connell, N. (1995). Large deviations and overflow probabilities for the general single-server queue, with applications. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge, UK: Cambridge University Press, vol. 118, pp. 363–374.
4. Elwalid, A. & Mitra, D. (1993). Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. *IEEE/ACM Transactions on Networking (TON)* 1(3): 329–343.
5. Gallier, J. (2012). Notes on convex sets, polytopes, polyhedra, combinatorial topology, Voronoi diagrams and Delaunay triangulations. Book in progress, http://www.cis.upenn.edu/jean/gbooks/convexpoly.html.
6. Ganesh, A., O'Connell, N., & Wischik, D. (2004). *Big queues*. Berlin, Heidelberg: Springer.
7. Glynn, P. & Whitt, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *Journal of Applied Probability* 31: 131–156.
8. Kella, O. (1993). Parallel and tandem fluid networks with dependent Lévy inputs. *The Annals of Applied Probability*, 3(3): 682–695.
9. Kesidis, G., Walrand, J., & Chang, C.-S. (1993). Effective bandwidths for multiclass Markov fluids and other ATM sources. *IEEE/ACM Transactions on Networking (TON)* 1(4): 424–428.
10. Kosiński, K. & Mandjes, M. (2015). Logarithmic asymptotics for multidimensional extremes under nonlinear scalings. *Journal of Applied Probability* 52(1): 68–81.
11. Mandjes, M. & Ridder, A. (1995). Finding the conjugate of Markov fluid processes. *Probability in the Engineering and Informational Sciences* 9(02): 297–315.
12. Rubino, G. & Tuffin B. (2009). *Rare event simulation using Monte Carlo methods*. Chichester, UK: John Wiley & Sons.
13. Seneta, E. (1973). *Non-negative matrices*. London: George Allen & Unwin.

APPENDIX

Proof of Theorem 3.1

Proof: This is shown by proving a number of implications.

∘ "2 ⇒ 1": Suppose there exist $i, j \in \mathcal{N}$ (possibly $i = j$) and $c, d \geq 0$ such that $cr_i^A + dr_j^A > 0$ and $cr_i^B + dr_j^B > 0$. Assume without loss of generality that a direct jump of the Markov process from state i to state j is possible. We may do this, because if this jump is not possible, we can jump from state i to state j in arbitrarily small time (because we assumed the Markov process is irreducible). Although this may have a very small probability, of importance is only that is has a *positive* probability. Let

$$T := \frac{(c+d)u}{cr_i^A + dr_j^A} \vee \frac{(c+d)u}{cr_i^B + dr_j^B}.$$

If $X_t = i$ for $0 \leq t < \frac{c}{c+d}T$ and $X_t = j$ for $\frac{c}{c+d}T \leq t \leq T$, then

$$A_T = \frac{1}{c+d}(cr_i^A + dr_j^A)T \geq u;$$

$$B_T = \frac{1}{c+d}(cr_i^B + dr_j^B)T \geq u.$$

∘ "3 ⇒ 2": This follows from Carathéodory's theorem for convex cones, see Gallier [5, Theorem 2.4], but we give a direct proof below. Define the sets PP, PM, MP and MM (P stands for plus and M for minus) as follows: $\text{PP} := \{k \in \mathcal{N} : r_k^A > 0, r_k^B > 0\}$, $\text{PM} := \{k \in \mathcal{N} : r_k^A > 0, r_k^B < 0\}$, etc. If there exists $k \in \text{PP}$, then we are done (choose $i = j = k$ and $c = d = 1$). Assume now without loss of generality that each $k \in \mathcal{N}$ is either in PM or MP. We can do this, because we can change the entries of $\vec{c}$ belonging to states in MM to zero, which makes both inner products even larger. Note now that neither PM or MP are empty. We let

$$i := \arg\max_{k \in \text{PM}} \frac{r_k^A}{-r_k^B}, \quad j := \arg\max_{k \in \text{MP}} \frac{r_k^B}{-r_k^A},$$

that is, i and j are the states which are the "closest" to the first quadrant in a geometrical sense. We now want to shift "mass" from the entries of $\vec{c}$ to c_i and c_j in such a way that both inner products do not decrease. A correct way of doing that is by constructing c, d as follows:

$$c := c_i + \sum_{k \in \text{PM}, k \neq i} c_k \frac{r_k^B}{r_i^B}, \quad d := c_j + \sum_{k \in \text{MP}, k \neq j} c_k \frac{r_k^A}{r_j^A}.$$

In this way, the negative contributions stay the same, while the positive contributions do not decrease. We have thus found $i, j \in \mathcal{N}$ and $c, d \geq 0$ with the desired requirements.

∘ "1 ⇒ 3": Suppose we have a sample path such that $A_t > u, B_t > u$ for some t. We can then write

$$A_t = \int_0^t r_{X_s}^A \, ds = \sum_{i=1}^{|\mathcal{N}|} c_i r_i^A, \quad B_t = \int_0^t r_{X_s}^B \, ds = \sum_{i=1}^{|\mathcal{N}|} c_i r_i^B,$$

for some $c_i \geq 0$. Choose $\vec{c} = (c_1, c_2, \ldots, c_{|\mathcal{N}|})$. ■